2008 Annual Report
Award#: N000140610565
PI: Matthew DeLisa, Cornell University

A. Scientific and Technical Objectives

Our long-term goal is to use tools from molecular biology to engineer multi-enzyme metabolic complexes, mimicking the physical forms ubiquitous in nature. The direct coupling between sequential enzymatic reactions, through either static or dynamic interactions, offers the promise of eliminating these production barriers as it reduces the distance between enzyme active sites and favors sequential reactions over diffusion into the bulk. Therefore, the objective of these studies is to engineer synthetic metabolic complexes by exploiting the assembly mechanisms of natural systems to spatially organize enzymes that participate in sequential reaction steps. We *expect* that by developing a generic set of tools to co-localize metabolic pathways, we will overcome traditional bottlenecks that limit the commercial viability of microbial factories. We have proposed the following specific aims:

(1) To demonstrate efficient multi-protein assembly in bacterial cells. To enable the intentional engineering of metabolic enzymes into functional metabolic complexes, we will explore a variety of novel methods for *in vivo* enzyme assembly.

(2) To Assemble functional metabolic complexes. We will utilize the intracellular assembly/cross-linking methodologies from Aim 1 to create synthetic metabolic complexes in bacteria that are capable of efficient metabolic conversions via fermentation of renewable resources. We have chosen as a model system the microbial synthesis of propylene glycol (1,2-propanediol or 1,2-PD).

(3) To enable combinatorial engineering of metabolic complexes via metabolite sensors. We will engineer a collection of intracellular switches that are capable of dynamically responding to intracellular metabolites (e.g., 1,2-PD) over a broad concentration range.

B. Approach

For Aims 1 and 2, our approach is to develop in vivo protein assembly/crosslinking strategies using modern molecular biology techniques. The first approach is direct translational fusion of the enzyme sequences resulting in a single covalently cross-linked fusion protein. The second approach is to employ a unique enzyme known as transglutaminase (TGase) that catalyze the post-translational modification of substrate proteins by the formation of covalent isopeptide bonds. A third approach is to graft “protein interacting domains” onto each pathway enzyme, thereby creating artificial interaction domains that promote intracellular enzyme assembly. These different assembly/cross-linking techniques will be developed in the context of the three-enzyme pathway that comprises 1,2-PD production in *Escherichia coli*. This entails cloning of the requisite constructs, expression in *E. coli* and examination of the 1,2-PD titers using HPLC/Mass Spec/NMR.

For Aim 3, our approach to develop metabolite sensors using protein-based switches that elicit a measurable activity upon small molecule binding. This is a departure from the original proposal as we have learned from experimentation that RNA-based switches are incapable of detecting small molecules that lack steric bulk. These protein-based switches will be useful for direct monitoring of intracellular 1,2-PD titers in living cells and are expected to open the door to laboratory evolution of our engineered metabolic complexes.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2008	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
2008 Annual Report	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Cornell University**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**SAR**	**14**	

C. Concise Accomplishments

The major accomplishments to date include:

(1) We have developed a spatial stochastic model of 1,2-PD biosynthesis (Conrado *et al.*, 2007) and demonstrated that compartmentalization of the three pathway enzymes comprising an engineered pathway for 1,2-PD biosynthesis leads to greatly improved kinetic properties for the pathway enzymes. We have recently employed a modified genome scale stoichiometric model of *E. coli* in conjunction with Dynamic Flux Balance Analysis (DFBA) in the presence and absence of crowding constraints to estimate the impact of cytosolic molecular crowding. A manuscript detailing these findings is in preparation.

(2) We have performed a thorough analysis of enzyme assembly mediated by (1) translational fusions; (2) protein interacting domains (PIDs) and (3) eukaryotic signaling scaffolds. To date, we have found that grafting of protein-protein interaction domains (PIDs) onto metabolic enzymes of interest results in the most significant improvement in 1,2-PD titers compared to cells expressing unassembled enzymes. This is significant as it confirms that enzyme organization is a powerful approach for developing highly efficient metabolic machinery in a recombinant organism. We have presented these findings in multiple venues (Conrado *et al.*, 2006; DeLisa, SIM 2006; Conrado & DeLisa, ACS 2007). In addition, a patent application has been disclosed and 2 manuscripts have been submitted.

(3) We have developed a chemical genetic reporter of protein stability that enables intracellular sensing of small compounds (Haitjema *et al.*, 2008). The significance of this tool is that it provides a fluorescence-based reporter of metabolite levels thereby opening the door to laboratory evolution of our metabolic 1,2-PD assemblies.

D. Expanded Accomplishments

D1. Simulation of 1,2-PD enzyme co-localization. Our first major accomplishment has been the demonstration that enzyme co-localization has a measurable effect on 1,2-PD titers in *Escherichia coli*. This was demonstrated via computer simulation in collaboration with Dr. Jeffrey Varner (Cornell University). Progress was made on these studies in year 1 of this grant; during year 2 we have performed an initial simulation study of the influence of molecular crowding on 1,2-PD production in *E. coli*. Unlike our previous stochastic modeling studies, we employed a modified genome scale stoichiometric model of *E. coli* in conjunction with Dynamic Flux Balance Analysis (DFBA) in the presence and absence of crowding constraints to estimate the impact of cytosolic crowding. We inserted the three-step 1,2-PD pathway (see Fig. 1) into the genome scale iJR904 *E. coli* stoichiometric model of Palsson

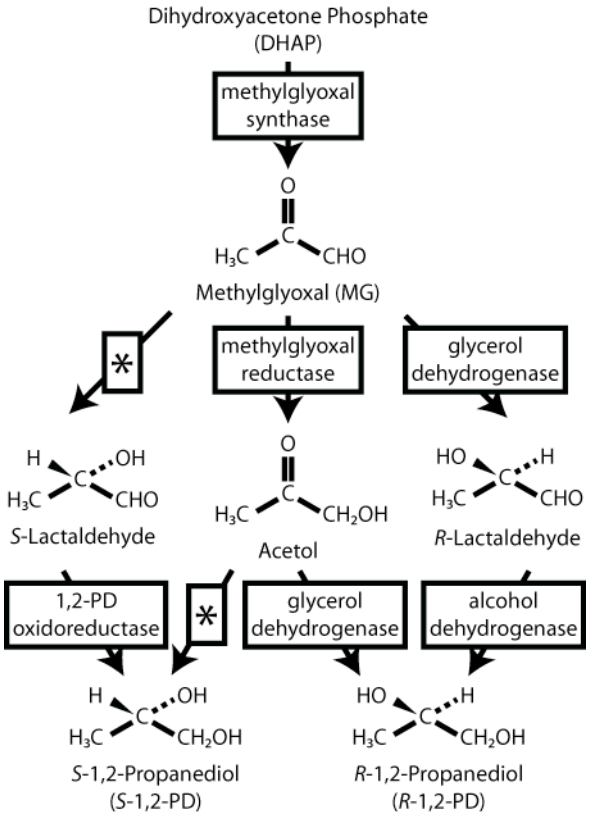

Figure 1. 1,2-PD metabolism. Arrows indicate an enzymatic reaction step, where names indicate the enzyme activity. A * indicates no activity has been isolated.

and coworkers (Reed *et al.*, 2003). The modified stoichiometric model was then used, in conjunction with a dFBA algorithm based upon previous work by Doyle and colleagues (Mahadevan *et al.*, 2002), to estimate batch and fed-batch anaerobic 1,2-PD fermentation trajectories with and without crowding constraints. Intracellular crowding was implemented as a single inequality constraint governing the aggregate flux through the modified *E. coli* network, similar to a recent simulation study (Beg *et al.*, 2007). Crowding significantly decreased the rate and yield of batch and fed-batch 1,2-PD production in a crowding coefficient specific manner (data not shown). Given that we know the *E. coli* cytoplasm is crowded, these proof-of-concept simulations suggest that overexpression of free enzymes in the 1,2-PD pathway may initially improve production, however, as the copy number of the pathway enzymes increases we may elicit biophysical crowding effects which will ultimately limit production. Second, the initial simulations suggested that rate and yield improvements in 1,2-PD production might be realized by reorganizing the *E. coli* cytoplasm using techniques such as molecular channeling (although we have not directly simulated the channeled case). A manuscript detailing these findings is in preparation.

Collectively, our simulation results suggest that enzyme co-localization is a powerful approach for improving the catalytic turnover of a channeled carbon substrate and should be particularly useful when applied to synthetic metabolic pathways that suffer from poor translation efficiency, are present in highly variable copy numbers, and have low turnover for new substrates. Furthermore, this approach represents a generic modeling framework for simultaneously analyzing spatial and stochastic events in cellular metabolism and should enable quantitative evaluation of the effect of enzyme compartmentalization on virtually any recombinant pathway.

D2. Enzyme compartmentalization towards production of 1,2-PD. In order to construct a synthetic metabolic pathway from DHAP to *R*-1,2-PD, we first needed to identify three sequential pathway enzymes (see Fig. 1). Since natural reaction pathways consume *R*- and *S*-lactaldehyde for reduction to lactate, a route through the acetol intermediate would suffer from less diverted flux. Along this path, the first and last enzymatic steps were well-defined, specifically the synthesis of MG by *E. coli* methylgloxal synthase (MgsA) and reduction of acetol by *E. coli* glycerol dehydrogenase (GldA) (Altaras & Cameron, 1999). However, the reduction of the intermediate, MG to acetol, can be performed by a number of NADH or NADPH dependent enzymes (Keseler *et al.*, 2005), most of which have not been tested *in vivo*. To test for activity on MG *in vivo*, we designed a high copy plasmid expressing MgsA, one of the following NADH (FucO, YdjG) or NADPH dependent enzymes (DkgA, DkgB), and GldA (Keseler et al., 2005). This was achieved in plasmid pBAD18, an arabinose inducible plasmid of pBR322 origin (Guzman *et al.*, 1995), where the three genes were translated from a polycistronic mRNA, each under control of separate but identical ribosome binding sites. This vector was used in all plasmids constructed below. Plasmids were transformed into wild-type *E. coli*, MC4100(Peters *et al.*, 2003), subcultured anaerobically in the presence of glucose, and extracellular levels of 1,2-PD were measured after fermentation by HPLC analysis (Altaras & Cameron, 1999). All genetic constructs produced significant levels of 1,2-PD with the NADH dependent enzymes showing higher activity towards MG demonstrated by higher 1,2-PD levels (Fig. 2). In moving forward, we

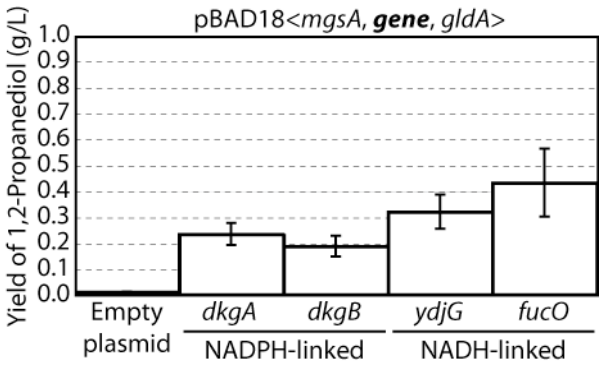

Figure 2. MG reductase enzymes. 1,2-PD production levels for NAD(P)H dependent enzymes coexpressed with *E. coli* MgsA and GldA within pBAD18.

chose FucO and DkgA, the highest NADH and NADPH dependent enzymes respectively.
D2.1. Fusion proteins. Since fusion proteins are a well-studied example (Conrado *et al.*, 2008), and are a simple test for analyzing the impact of enzyme compartmentalization, our first approach began with engineering our pathway enzymes, covalently attaching 2 enzymes by encoding both active sites on the same polypeptide (Fig. 3a). In this way, the active sites could be brought into close proximity and therefore allow for the proposed synergistic activity found in natural systems, like the tryptophan synthase channel (Conrado *et al.*, 2008). In order to systematically test this approach, the two design elements considered were: (1) the linker length and composition; and (2) the order of the protein fusions within the context of the entire pathway. Despite a large body of knowledge regarding natural linkers within multidomain proteins (Argos, 1990), there is little consensus on the length and composition of synthetic linkers when connecting two proteins that are normally not fused (Conrado et al., 2008). In order to find a suitable candidate, we systematically applied a set of six well-studied linker peptides used for *E. coli* fusion proteins, which varied in composition and length, from 5 to 37 amino acids (Table 1) (Chang et al., 2005). This approach allowed us to test the role of the linker in terms of its effect on protein stability, as well as the effect of linker length on the degree of synergistic coupling, i.e., if the kinetic benefit disappeared with longer linkers due to the increased diffusional distance. This resulted in several in-frame genetic fusions between the first two enzymatic steps of the pathway, denoted *mgsA*-L-*dkgA* or *mgsA*-L-*fucO* where L represents any one of the linkers, with the third enzyme freely expressed in a bicistronic message. As shown in Figure 4, a 2-3 fold increase in 1,2-PD production resulted from many of the enzyme fusions, consistent with *in vitro* studies on fusion proteins (Conrado et al., 2008). In the case of the *dkgA* fusion, the increase in activity was seen regardless of linker, however a similar fold improvement was only observed with the longer linkers in the *fucO* fusions. These promising results motivated us to determine whether or not a similar benefit would be seen when fusing only the second two pathway enzymes, thus channeling acetol towards 1,2-PD. Applying the same set of linkers, the following genetic fusions were created: *dkgA*-L-*gldA* and *fucO*-L-*gldA*, and with free *mgsA* encoded bicistronically. Interestingly, this set of fusion constructs (Fig. 5) resulted in a slight decrease in yield, regardless of linker or pathway enzyme, suggesting that no kinetic benefit was observable and/or that these fusions were not well tolerated.

Table 1. Peptide Linkers and their Properties (reproduced from (Chang et al., 2005)

Name	Sequence	Properties
L5	TSAAA	5-aa linker that results from consecutive SpeI and NotI restriction endonuclease sites
L15a	TSMTATADVLAMAAA	15-aa naturally occurring inter-domain linker highly conserved across kingdoms
L15b	TSGGSGGSGGSGAAA	15-aa uncharged flexible linker previously used to fuse multi-domain proteins
L16	TSGSAASAAGAGEAAA	16-aa flexible linker previously used in combination with GFP in *E. coli*
L25	$TS(GGGGS)_4AAA$	25-aa flexible linker used extensively in recombinant antibody fragments
L37	$TSAG(EAAAK)_6AAA$	37-aa α-helical linker used for intramolecular FRET between fluorescent proteins

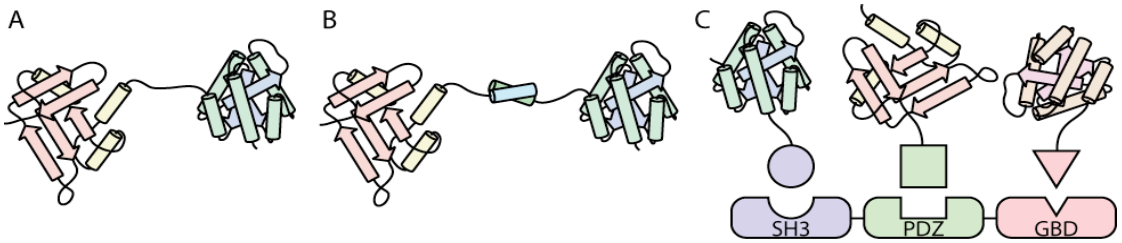

Figure 3. Engineering enzyme compartmentalization. (A) Fusion proteins connected by a peptide linker, (B) Chimeric proteins fused to leucine zippers connected by post-translational assembly of PIDs, (C) Post-translational assembly of enzyme complexes where the SH3-PDZ-GBD fusion represents the three specific docking sites for each enzyme-ligand fusion.

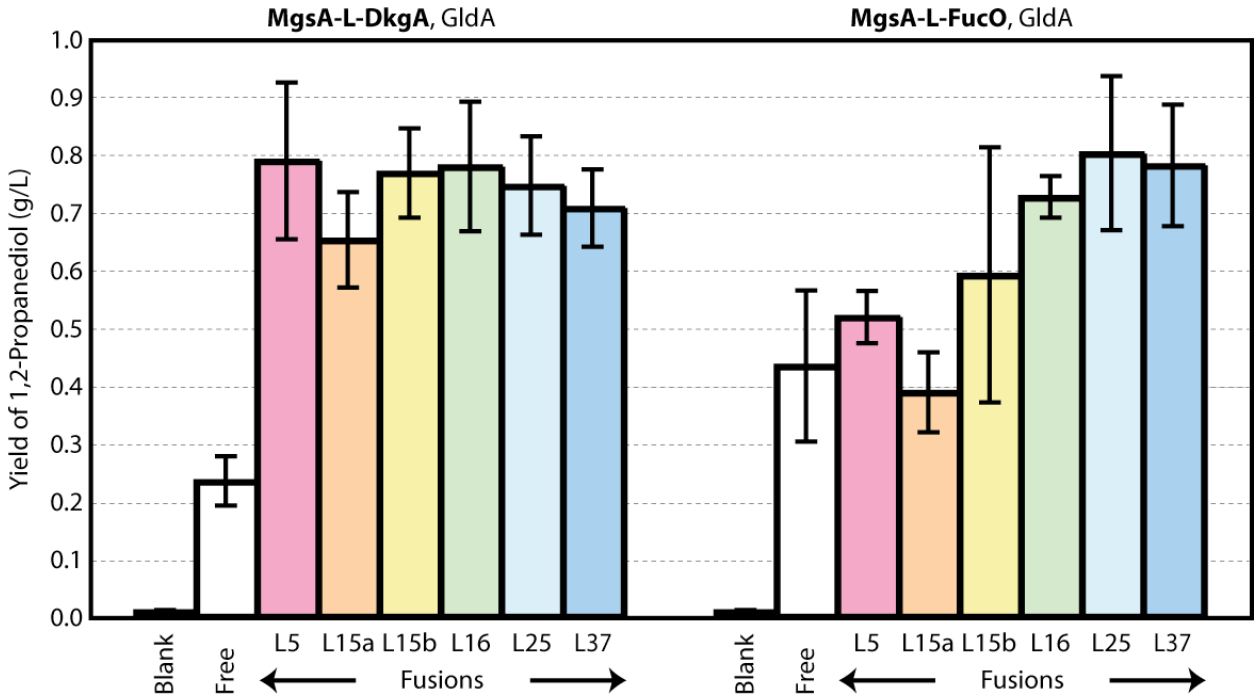

Figure 4. Fusion proteins between 1st and 2nd pathway enzymes. 1,2-PD production levels for several MgsA-DkgA or MgsA-FucO fusions coexpressed with GldA, using linkers listed in **Table 1**. Also shown are the empty plasmid (blank) and free enzyme controls. All bars represent the average of three experiments with error bars representing the standard deviation.

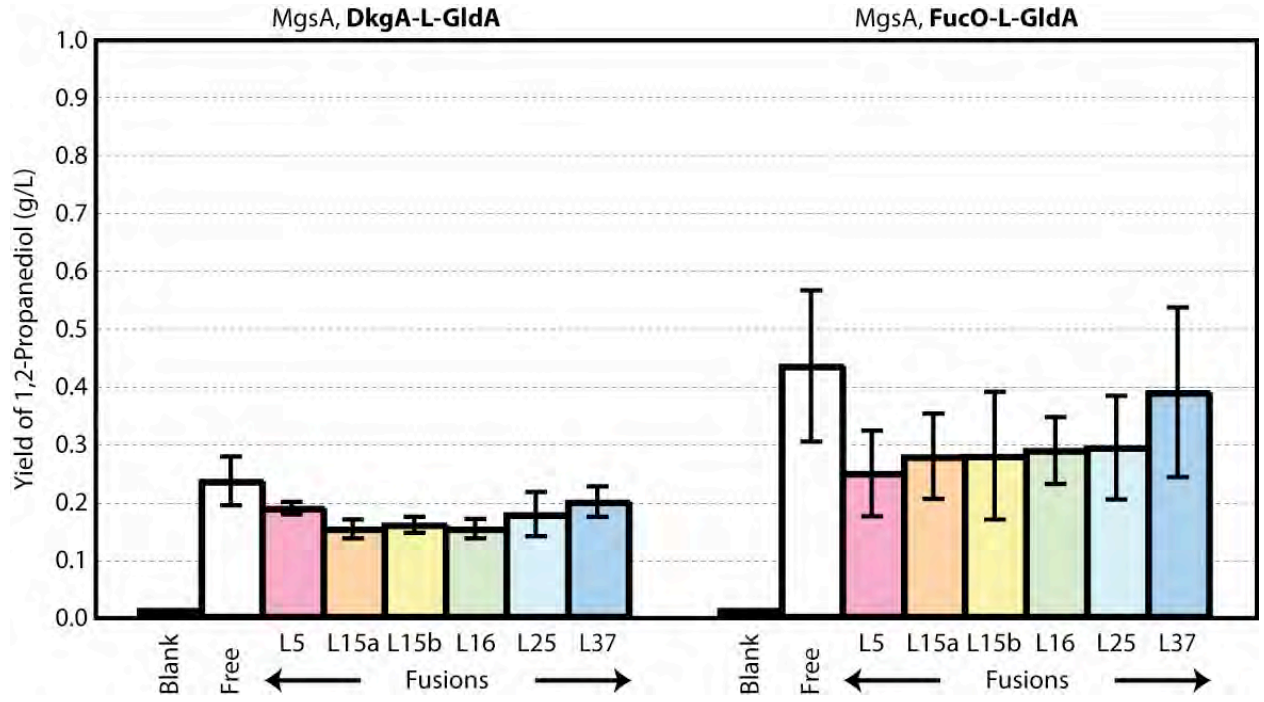

Figure 5. Fusion proteins between 2nd and 3rd pathway enzymes. 1,2-PD production levels for several DkgA-GldA or FucO-GldA fusions coexpressed with MgsA, using linkers listed in **Table 1**. Also shown are the empty plasmid (blank) and free enzyme controls. All bars represent the average of three experiments with error bars representing the standard deviation.

As a final goal of compartmentalizing a set of sequential pathway enzymes, we decided to move ahead with a three-gene fusion, to determine if a fold improvement could be detected over the *mgsA*-L-*dkgA*/*fucO* fusions due to altered kinetics over the free enzymes (Fig. 4). Selecting the successful linkers in both cases, the L16 linker was used between the first two pathway enzymes and the L37 linker between the last two enzymes, thus making MgsA-L16-DkgA/FucO-L37-GldA fusions. After being constructed, these were compared to the corresponding free and fusion enzyme systems by their production of 1,2-PD (Fig. 6). Combining the fusions between all pathway genes did not yield an improvement over the simple fusion of the first two enzymes. This was not altogether surprising since the fusion of the last two enzymes seemed to impair enzyme activity from the decrease in 1,2-PD production above (Fig. 5). In probing for the explanation for these disparate levels of production, a Western Blot was performed to analyze the intracellular protein levels, which might be impacted from these novel fusions. In comparing the protein levels of the second pathway enzyme to each of selected fusions, there are significant variations in protein levels in both the soluble and insoluble fractions (Fig. 6). The MgsA-L16 fusions show a dramatic increase in 1,2-PD production, and interestingly there is also an increase in the protein levels of these samples, which both appear in the soluble and insoluble fractions. Likewise the L37 fusions show a slight decrease in protein levels, which translates to a decrease in 1,2-PD production. Despite the enzyme compartmentalization that might occur here, acetol is not believed to be a branch point substrate, and therefore its channeling has little effect on 1,2-PD production. When moving toward the 3-enzyme fusion, we see both degradation of the polypeptide chain, in the case of DkgA, as well as a sharp decrease in protein level, in the case of FucO. Here it seems that the little soluble protein is active, but is not present in large quantities due to problems with folding and aggregation that result from non-natural fusions. At this point it is important to note that each of the pathway enzymes is multimeric, and the active structures are denoted $MgsA_6$, $DkgA_2$, $FucO_2$, and $GldA_8$. When combining these polypeptides by gene fusion, especially when combining more than two genes as above, it has important consequences in terms of aggregation and achieving active enzyme units (Fig. 7). While aggregation may be a general problem with compartmentalizing multimeric enzymes, it is important to pursue alternative strategies that might allow these to form in a manner that they maintain activity.

D2.2. Protein Interacting Domains (PIDs). In moving beyond protein fusions as a result of the fusion instability when colocalizing more than 2 enzymes, we wanted to couple the enzyme active sites in a way that would allow proper protein folding, and possibly subunit assembly, before compartmentalizing the pathway enzymes. Similar to the assembly of metabolic channels like the tryptophan synthase $\alpha\beta\beta\alpha$ complex (Conrado *et al.*, 2008), a post-translational, non-covalent assembly of sequential pathway enzymes would likely allow for more stable enzyme formation of active multimers. To achieve this, we employed known protein interacting domains, that when fused N- or C-terminally to our pathway enzymes, would bind together and would thus bring the sequential active sites into close proximity (Fig. 3b). This strategy offers more design flexibility than gene fusion above, because in addition to design of the linker region, there are also considerations for the binding associations, homo- or heterodimerization, and their affinity in terms of the K_D. Here our design criteria were domains that: (1) are well expressed in *E. coli*, (2) possess high affinity interactions in the μM to sub-μM range, given that natural proteins can reach 1-10 μM concentrations in *E. coli* (Sceller *et al.*, 2000), (3) are short in length, preferably <50 amino acids, (4) are characterized by highly specific interactions with little cross-reactivity. Based on these standards, we selected 3 sets of interacting domains for further analysis (Table 2).

Table 2. Protein Interacting Domains and their Properties

Name	Type	Association Type	Size (AAs)	Affinity - K_D (μM)	Refs.
GCN4/GCN4	Leucine Zipper	Homodimerization	48 / 48	0.5	(Dragan *et al.*, 2004)
cJun/cFos	Leucine Zipper	Heterodimerization	41 / 40	0.001-0.11	(Pernelle *et al.*, 1993, Patel *et al.*, 1994, Oyama *et al.*, 2006, Heuer *et al.*, 1996)
SH3/ SH3ligand	Protein Interacting Domains	Heterodimerization	57 / 11,8	0.1, 10	(Dueber *et al.*, 2007)

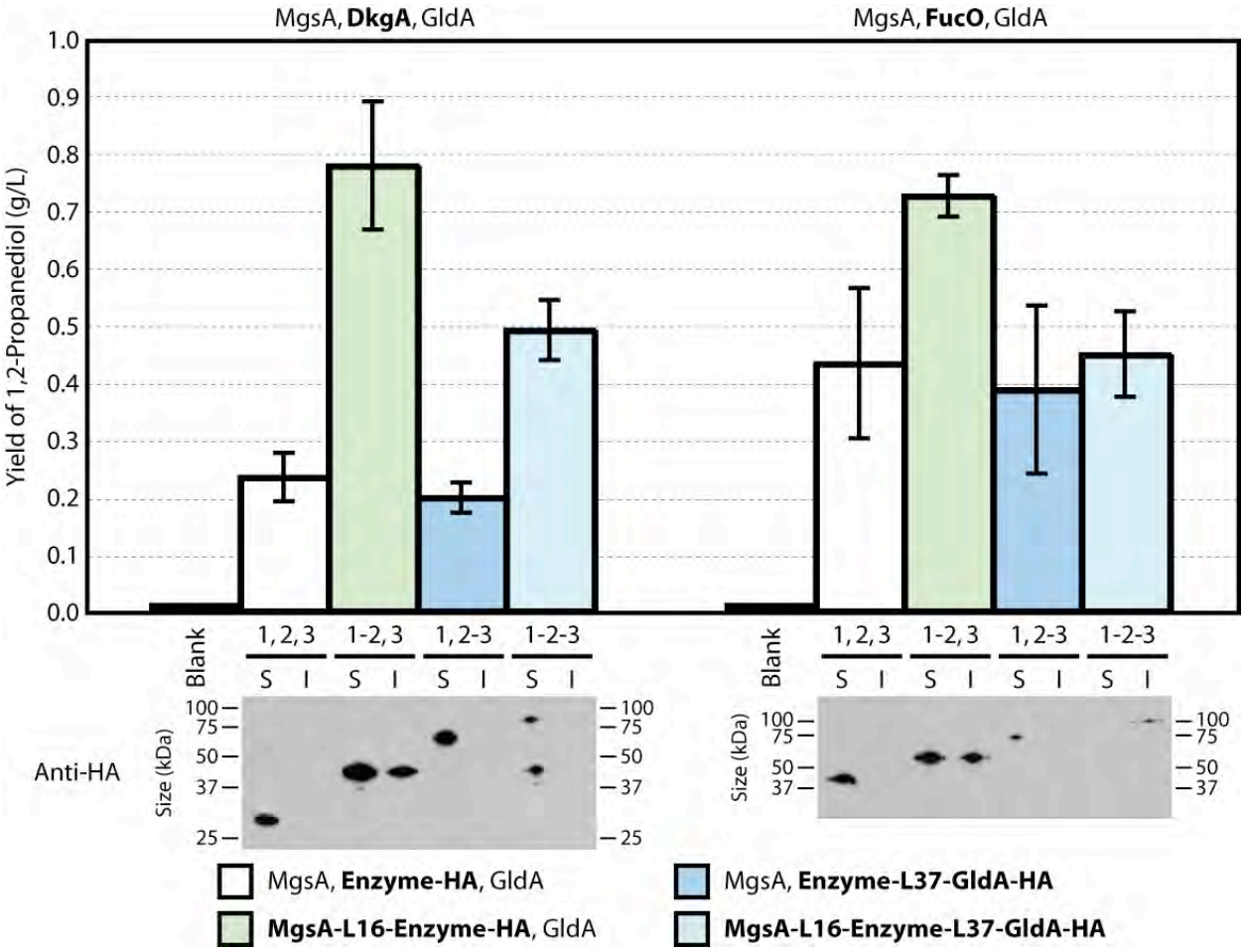

Figure 6. Compartmentalization by fusion proteins. 1,2-PD production levels for empty plasmid, free enzyme (1,2,3), and each fusion construct using a single linker or both. Western blots of soluble (S) and insoluble (I) cell fractions against Anti-HA antibody. All bars represent the average of three experiments with error bars representing the standard deviation.

These domains were fused to each of the first two pathway enzymes with the L16 linker in a polycistronic mRNA with the third enzyme freely expressed (e.g., MgsA-L16-cJun; cFos-L16-DkgA; GldA). In this manner, we were able to measure the degree of synergistic effect by comparison to a polycistronic free enzyme system. Additionally, the PIDs offer further controls in that in each of these interacting domains, (1) point mutations can be made to reduce binding affinity and (2) cross-reacting species can be combined (e.g., MgsA-SH3; cFos-DkgA). These additional controls should help to elucidate the mechanism of the fold-improvement seen in protein fusions, namely whether the increase was due to enzyme compartmentalization or an increase in protein stability as seen by Western blot.

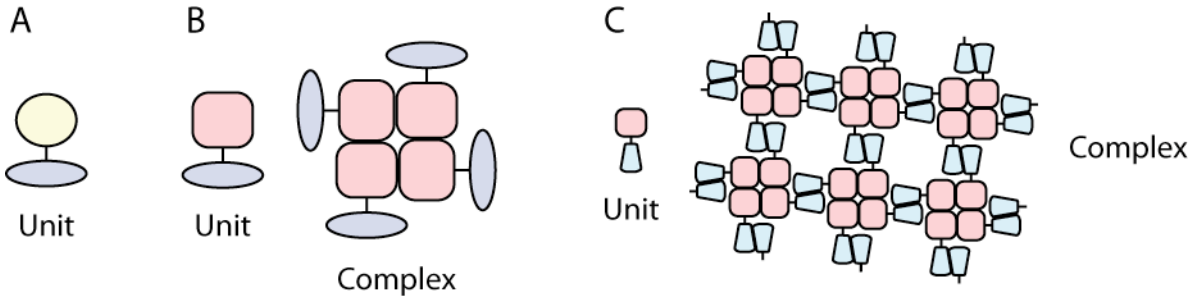

Figure 7. Complexes of multimeric fusion proteins. Fusion protein between (a) two monomers, (b) a monomer and a tetramer, (c) a dimer and a tetramer, forming a multimeric protein aggregate in this last case (adapted from (Bulow & Mosbach, 1991)).

These constructs were analyzed as above for their extracellular 1,2-PD production levels following fermentation, shown in Figure 8, with each of the interacting domains and their inactivating control. Interestingly for the DkgA or FucO system, the GCN4 leucine zippers were responsible for the largest increase in production, followed by either cJun/cFos or the SH3/SH3ligand pair. For the FucO system, the GCN4 PIDs resulted in the largest increase observed thus far, exceeding the fusion proteins. Looking towards the controls for each of these interacting domains, the inactivating mutations to GCN4 (Leder *et al.*, 1995), had a strong decrease in 1,2-PD levels, though still higher than free enzyme control. This trend was similar with the cJun/cFos control (Ransone *et al.*, 1989), although the weaker SH3 ligand (Dueber et al., 2007) gave similar production to that of the strong ligand. However, the Western Blots on these samples were particularly telling in revealing the effect of enzyme expression with these novel fusions (Fig. 8). Interestingly, the GCN4 fusions significantly increased the stability of enzymes. Here we see some degradation at the L16 linker, especially with the proline mutations, which explains the decrease in activity of these samples. The cJun/cFos and the SH3 fusions had significantly reduced protein levels, despite their relatively high activity. This suggests that enzyme compartmentalization plays a strong role in these cases. Additionally, although the mutations in these last 2 cases reduce the affinity 100-fold, the K_D values may still be below the intracellular enzyme concentrations, which might allow channeling to be active here. Based on these encouraging results, we constructed a 3-enzyme interacting system (denoted MgsA-GCN4; GCN4-DkgA/FucO-cJun; cFos-GldA). Despite the promise of this system, our initial results yielded lower 1,2-PD levels than the original GCN4 system (results not shown). This mimics the trend seen when moving from 2 to 3-enzyme compartmentalization with fusion proteins (Fig. 6). As in the case of the fusion enzymes, it is believed here that the middle enzyme is constrained in folding because of the presence of both GCN4 and cFos domains at the N- and C-termini, respectively.

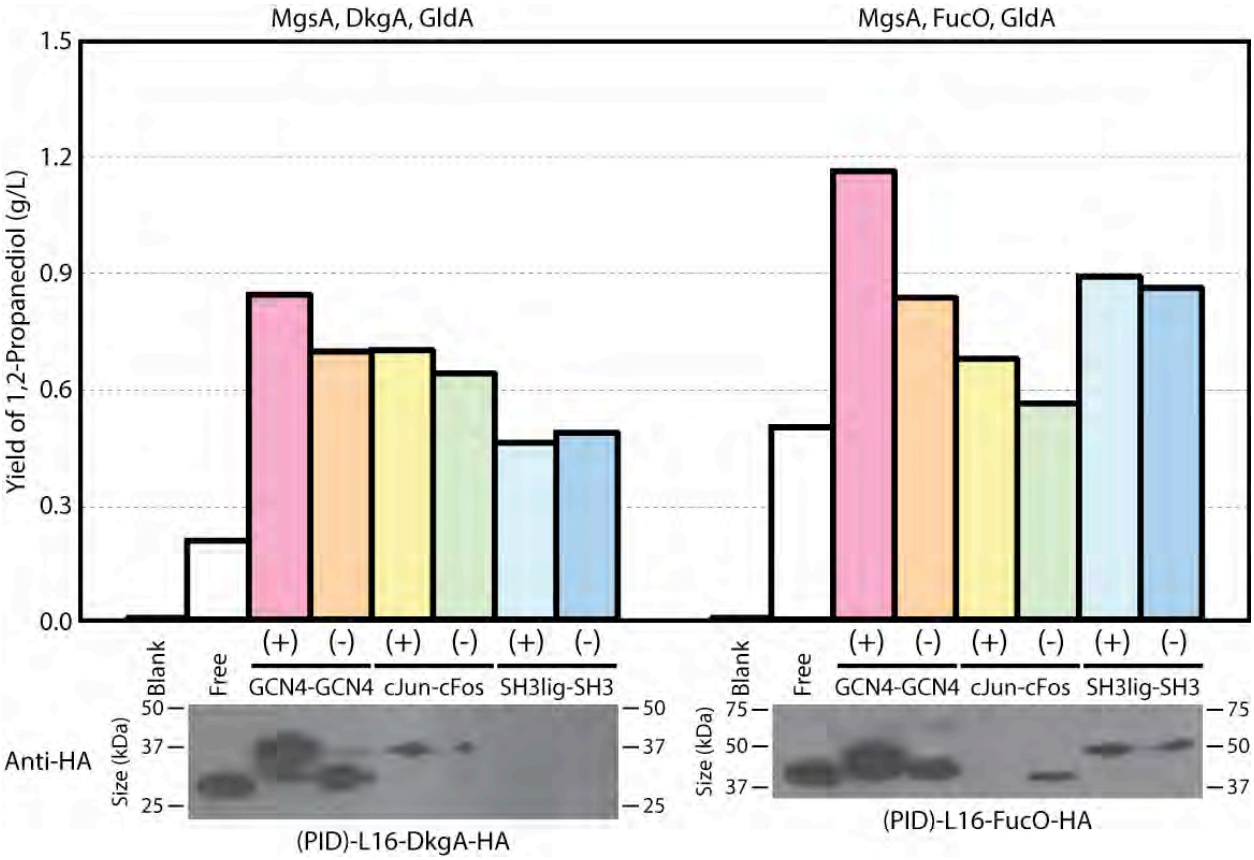

Figure 8. Compartmentalization using PIDs. 1,2-PD production using several PID systems, listed in Table 2, as well as the empty plasmid and free enzyme control. PIDs are fused C-terminally to MgsA and N-terminally to DkgA or FucO. GldA was coexpressed. Western blots of soluble fraction against Anti-HA antibody. (+) indicates wild-type PIDs where (-) indicates PIDs with inactivating mutation.

D3. Chemical genetic control of protein stability results in an intracellular sensor of small molecules. Since there are currently no generic reporters for intracellular metabolites, we sought to develop a tool for sensing such compounds. Our approach was to develop a protein conformational switch comprised of an unstable domain and a reporter protein (Fig. 9). The unstable domain was selected such that introduction of a small molecule ligand that stabilizes the domain would restore stability of the entire fusion and thus lead to measurable activity of the reporter protein. The reporter for our switch was chosen as the green fluorescent protein (GFP) so that upon introduction of small molecules that stabilize the unstable domain, a large increase in cell fluorescence would result (Fig. 10a). For the unstable domain, we chose the TraR transcriptional activator from *Agrobacterium tumefaciens* (Fig. 10a). In the absence of its natural ligand, the freely diffusible quorum signaling molecule 3-oxooctanyl-l-homoserine lactone (OHHL), the TraR protein is a monomer that is highly unstable in the cytoplasm of

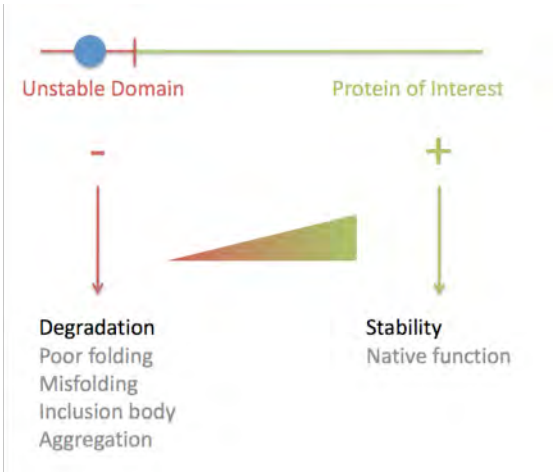

Figure 9. Chemical genetic control of protein stability.

Agrobacteria tumefaciens and *E. coli* (Zhu & Winans, 2001). However, upon biding of OHHL, TraR forms an extremely stable dimer (Vannini *et al.*, 2002). We have observed this same OHHL-dependent stability with an engineered TraR-GFP fusion protein. That is, in the absence of OHHL, TraR-GFP is highly unstable and cells expressing the fusion are relatively non-fluorescent (Fig. 10b). However, upon addition of OHHL, the TraR-GFP protein is stabilized (presumably in a dimeric conformation) and the cells become highly fluorescent in an OHHL dose-dependent fashion (Fig. 10b). This work was recently submitted for publication (Haitjema et al., 2008). We are now exploring the further engineering of this GFP-TraR conformational switch for sensing molecules other than OHHL. We expect that a collection of small molecule switches can be created using the GFP-TraR backbone, simply by the application of protein design and/or laboratory evolution to change the substrate specificity of TraR from OHHL to other compounds of interest such as D-BT.

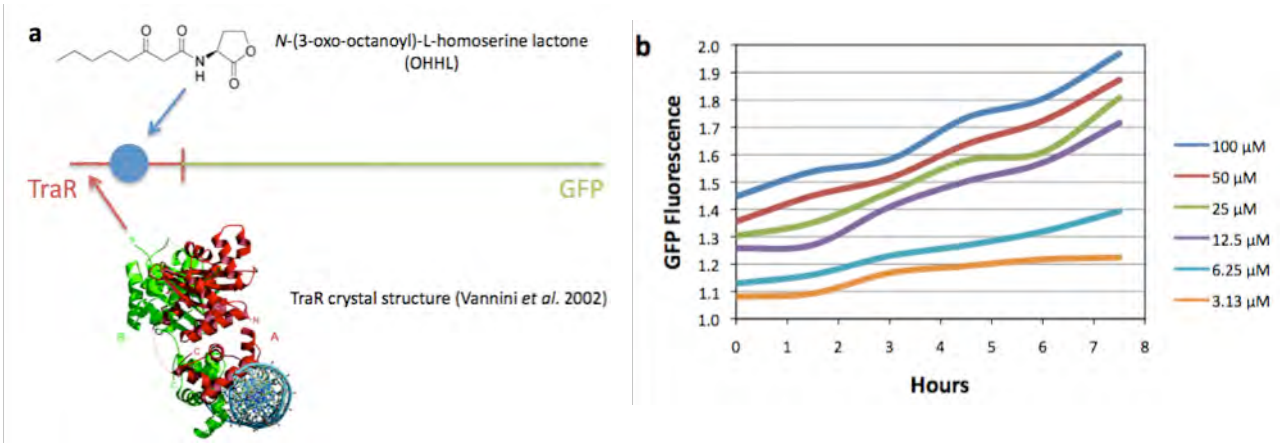

Figure 10. (a) TraR-GFP fusion as a reporter of small molecules in the cytoplasm of *E. coli* cells. (b) Dose-dependent response of the TraR-GFP-expressing *E. coli* to various concentrations of OHHL added exogenously to the growth medium. Cells were grown in 96-well plates and assayed using a fluorescent plate reader.

E. Work Plan

E1. Simulation studies. Looking forward, we will validate our initial simulation studies by conducting a series of continuous culture experiments using 1,2-PD producing *E. coli* variants in which the model will be used to estimate measurable physiological parameters (specific glucose uptake and 1,2-PD production rates as well as the specific rates of by-product formation) as a function of crowding. We will then compare the observed and estimated physiological rates as a means of estimating the *E. coli* crowding coefficient. We expect these experiments will support our contention that *E. coli* has a crowded cytoplasmic environment and will underscore the need to spatially organize the 1,2-PD pathway.

To fully realize the benefit of metabolic channeling, we must maximize the flow of 1,2-PD precursors into the assembly, i.e., we must embed the 1,2-PD assembly into a metabolic strain background that has been engineered to take advantage of the benefits offered by channeling. To design optimal precursor flux to the 1,2-PD channel, we will develop new network design tools with assistance from our collaborator Dr. Jeffrey Varner (Cornell University) that can be used to computationally develop metabolic architectures that take full advantage of engineered assemblies. Our strategy will use error tolerant kinetic models, based upon the cybernetic modeling paradigm, in conjunction with a non-linear programming procedure inspired by

OPTKNOCK to calculate optimal network configurations that maximize flux to 1,2-PD assemblies while simultaneously satisfying design and process constraints.

E2. New approaches to protein assembly

E2.1 Protein Scaffolds. While showing success in increasing the yields of synthetic chemicals for 2-enzyme channels, fusion proteins and interacting domains fall short when trying to compartmentalize entire metabolic pathways. The problem seems to hinge at the protein instability that results from large fusions at both the N- and C-termini of the native protein sequence. To overcome this, we will create a stable scaffold *in vivo* with several highly specific and modular docking sites to which we can successful compartmentalize successive pathway enzymes for enhanced metabolite production. To this extent, we will begin with the chimeric protein scaffold, composed of various eukaryotic protein domains (Table 3), that has been tested both *in vitro* by others (Dueber et al., 2007) and in our lab (Fig. 11). The development of this scaffold would provide a generic framework for colocalization of other metabolic pathways in that the required fusion to the sequential enzymes is short, and not likely to interfere with protein folding. Here, our pathway enzymes would colocalize on this scaffold by several short ligands, which specifically target each of these docking domains. Using the SH3-PDZ-GBD scaffold construct from this work, the corresponding ligands can be fused N- or C-terminally using our optimal linker and previous orientation studies. In the metabolic pathway towards 1,2-propanediol (1,2-PD), MgsA will be targeted to the SH3 domain, FucO to the PDZ domain, and GldA to the GBD domain. For the longer pathway towards D-1,2,4-Butanetriol (BT), the final 3 enzymes will be targeted to the scaffold since the final enzymatic steps suffer from several competing reactions. Here YjhG will be targeted to SH3, MdlC to the PDZ domain, and AdhP to the GBD domain. To accomplish this, we will fuse the strong ligands (listed in Table 3) C-terminally to the pathway enzymes by means of the L16 linker to target each to the protein scaffold.

E2.2 DNA Scaffolds. As an alternative to colocalizing metabolic pathways along a protein scaffold, we will also develop a novel DNA-based scaffold which might prove to be more abundant, stable, and amenable to alteration (Fig. 12). Here a DNA-based scaffold is an attractive alternative as these stable genetic elements (e.g. plasmid DNA) are well defined and easily modified, and further, there are many known DNA binding proteins that can target our pathway enzymes along the DNA platform. Specifically a number of zinc finger domains have been evolved for high specificity towards 49 of the 64 three-base pair (bp) recognition sequences. These have been genetically combined to create multidomain proteins capable of targeting 18 base pair sites that can discriminate single base pair changes with up to 100-fold loss in affinity (Mandell & Barbas, 2006, Segal *et al.*, 1999, Liu *et al.*, 1997). We will employ this technique to create promoterless high-copy plasmids that encode interspaced docking sites, here only 12 base pairs in length. Each zinc finger domain, 22 amino acids in length, targets 3 sequential DNA base pairs; therefore we will fuse 4 zinc finger elements together to create motif capable to specifically targeting our plasmid docking sites, without suffering from competition of genomic DNA or size instability. This technique will enable the co-localization of longer metabolic pathways, e.g. D-BT biosynthesis, and will enable further engineering improvements by controlling the spacing of the sequential enzymes and their geometric arrangements. This will enable multiple repeats of slow enzymes within a colocalized element, and also several repeats of the entire pathway along a single plasmid DNA *in vivo* to along for toggling the number of channels.

Table 3. Protein Scaffold Components and their Properties (reproduced from Dueber et al., 2007)

Docking Domain	Source	Ligand (AAs)	K_D (μM)
SH3	Mouse Crk (134-191)	PPPALPPKRRR (11) PPPVPPRR (8)	0.1 10
PDZ	Mouse a-syntrophin (77-171)	GVKESLV (7) GVKQSLL (7)	8 100
GBD	Rat N-WASP (196-274)	SGIVGALMEVMQKRSKAIH (19)	1

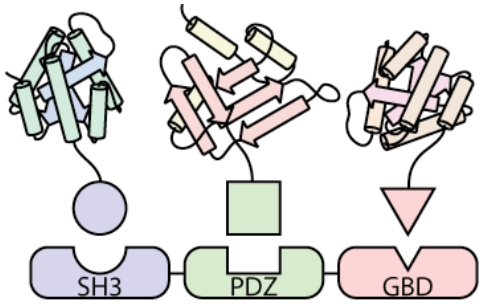

Figure 11. Protein scaffold consisting of 3 fused, eukaryotic protein domains, SH3, PDZ, and GBD. Metabolic enzymes for 1,2-PD or BT biosynthesis are colocalized to the scaffold by strong binding ligands (Table 3).

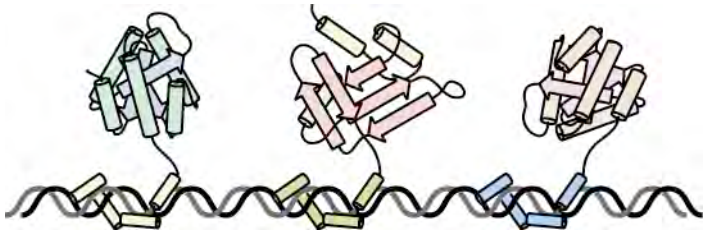

Figure 12. DNA scaffold from a plasmid, showing the protein-binding region, spaced 12 base pair elements. Metabolic enzymes are colocalized along the DNA by zinc finger domains fused C-terminally.

E3. Chemical genetic sensors of 1,2-PD. Our results to date have clearly demonstrated that the TraR-GFP fusion has an exquisite ability to be stabilized by the binding of a small molecule ligand and thus "sense" the presence of extremely small compounds. Moving forward, we will focus on the development of TraR-GFP sensors that respond to an array of different compounds. This will entail the use of rational design and laboratory evolution in order to alter the substrate specificity of TraR-GFP. We will initially focus on the substrate 1,2-PD.

F. Major Problems/Issues. There have been no significant problems.

G. Technology Transfer. We have disclosed one new invention during this period.

Conrado, R.J. and DeLisa, M.P. "Compositions and methods for intracellular enzyme assembly and uses thereof" invention disclosed.

References Cited

Altaras, N. E. & D. C. Cameron, (1999) Metabolic engineering of a 1,2-propanediol pathway in Escherichia coli. *Appl Environ Microbiol* **65**: 1180-1185.

Argos, P., (1990) An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *J Mol Biol* **211**: 943-958.

Beg, Q. K., A. Vazquez, J. Ernst, M. A. de Menezes, Z. Bar-Joseph, A. L. Barabasi & Z. N. Oltvai, (2007) Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. *Proc Natl Acad Sci U S A* **104**: 12663-12668.

Bulow, L. & K. Mosbach, (1991) Multienzyme systems obtained by gene fusion. *Trends Biotechnol* **9**: 226-231.

Chang, H. C., C. M. Kaiser, F. U. Hartl & J. M. Barral, (2005) De novo folding of GFP fusion proteins: high efficiency in eukaryotes but not in bacteria. *J Mol Biol* **353**: 397-409.

Conrado, R. J., T. J. Mansell, J. D. Varner & M. P. DeLisa, (2007) Stochastic reaction-diffusion simulation of enzyme compartmentalization reveals improved catalytic efficiency for a synthetic metabolic pathway. *Metab Eng* **9**: 355-363.

Conrado, R. J., J. D. Varner & M. P. DeLisa, (2008) Engineering the compartmentalizaion of metabolic enzymes: mimicking nature's synergy. *Current Opinion in Biotechnology*.

Dragan, A. I., L. Frank, Y. Y. Liu, E. N. Makeyeva, C. Crane-Robinson & P. L. Privalov, (2004) Thermodynamic signature of GCN4-bZIP binding to DNA indicates the role of water in discriminating between the AP-1 and ATF/CREB sites. *Journal of Molecular Biology* **343**: 865-878.

Dueber, J. E., E. A. Mirsky & W. A. Lim, (2007) Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat Biotechnol* **25**: 660-662.

Guzman, L. M., D. Belin, M. J. Carson & J. Beckwith, (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**: 4121-4130.

Haitjema, C., A. C. Fisher & M. P. DeLisa, (2008) Chemical genetic control of protein stability results in an intracellular sensor of small molecules. *J Amer Chem Soc (submitted)*.

Heuer, K. H., J. P. Mackay, P. Podzebenko, N. P. S. Bains, A. S. Weiss, G. F. King & S. B. EasterbrookSmith, (1996) Development of a sensitive peptide-based immunoassay: Application to detection of the Jun and Fos oncoproteins. *Biochemistry* **35**: 9069-9075.

Keseler, I. M., J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I. T. Paulsen, M. Peralta-Gil & P. D. Karp, (2005) EcoCyc: a comprehensive database resource for Escherichia coli. *Nucleic Acids Res* **33**: D334-337.

Leder, L., C. Berger, S. Bornhauser, H. Wendt, F. Ackermann, I. Jelesarov & H. R. Bosshard, (1995) Spectroscopic, calorimetric, and kinetic demonstration of conformational adaptation in peptide antibody recognition. *Biochemistry* **34**: 16509-16518.

Liu, Q., D. J. Segal, J. B. Ghiara & C. F. Barbas, 3rd, (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci U S A* **94**: 5525-5530.

Mahadevan, R., J. S. Edwards & F. J. Doyle, 3rd, (2002) Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophys J* **83**: 1331-1340.

Mandell, J. G. & C. F. Barbas, 3rd, (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* **34**: W516-523.

Oyama, R., H. Takashima, M. Yonezawa, N. Doi, E. Miyamoto-Sato, M. Kinjo & H. Yanagawa, (2006) Protein-protein interaction analysis by C-terminally specific fluorescence labeling and fluorescence cross-correlation spectroscopy. *Nucleic Acids Research* **34**: -.

Patel, L. R., T. Curran & T. K. Kerppola, (1994) Energy-Transfer Analysis of Fos-Jun Dimerization and DNA-Binding. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 7360-7364.

Pernelle, C., F. F. Clerc, C. Dureuil, L. Bracco & B. Tocque, (1993) An Efficient Screening Assay for the Rapid and Precise Determination of Affinities between Leucine-Zipper Domains. *Biochemistry* **32**: 11682-11687.

Peters, J. E., T. E. Thate & N. L. Craig, (2003) Definition of the Escherichia coli MC4100 genome by use of a DNA array. *Journal of Bacteriology* **185**: 2017-2021.

Ransone, L. J., J. Visvader, P. Sassone-Corsi & I. M. Verma, (1989) Fos-Jun interaction: mutational analysis of the leucine zipper domain of both proteins. *Genes Dev* **3**: 770-781.

Reed, J. L., T. D. Vo, C. H. Schilling & B. O. Palsson, (2003) An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biol* **4**: R54.

Sceller, L. L., C. Ripoll, M. Demarty, A. Cabin-Flamand, T. Nystrom, M. S. Jnr. & V. Norris, (2000) Modelling bacterial hyperstructures with cellular automata. *Interjournal Paper 366*.

Segal, D. J., B. Dreier, R. R. Beerli & C. F. Barbas, 3rd, (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A* **96**: 2758-2763.

Vannini, A., C. Volpari, C. Gargioli, E. Muraglia, R. Cortese, R. De Francesco, P. Neddermann & S. D. Marco, (2002) The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. *EMBO J* **21**: 4393-4401.

Zhu, J. & S. C. Winans, (2001) The quorum-sensing transcriptional regulator TraR requires its cognate signaling ligand for protein folding, protease resistance, and dimerization. *Proc Natl Acad Sci U S A* **98**: 1507-1512.